Transcript Pooling or "TraP"

A Comprehensive Method for Platform Performance Evaluation

David Gerhold

Merck & Co.

Platform Performance Evaluation

- Reproducibility
  - How many replicate hybridizations do I need to detect gene regulation events?
  - How many replicate hybridizations do I need to eliminate false-positive events?
- Sensitivity
  - What is the detection limit for each gene?
  - Which oligos are most sensitive?!
- Specificity
  - Can I distinguish gene family members?
  - Which oligos are gene-specific?!

Approaches to Evaluating Microarray Platforms for Sensitivity, Specificity, & Reproducibility

- Hybridize two RNA populations from biological sources. Such as liver *vs*. brain, & liver *vs*. (same) liver.
  - Assess: reproducibility, sensitivity, & specificity

- Make synthetic DNA/RNA oligonucleotides complementary to (all?) probes/cDNAs on array. End label and hybridize. Stable but $$.
  - Assess: reproducibility, sensitivity, & specificity

  for each probe/cDNA

- Label and hybridize *in vitro* transcripts for many genes on the array.
  - Assess: reproducibility, sensitivity, & specificity

  for each probe/cDNA

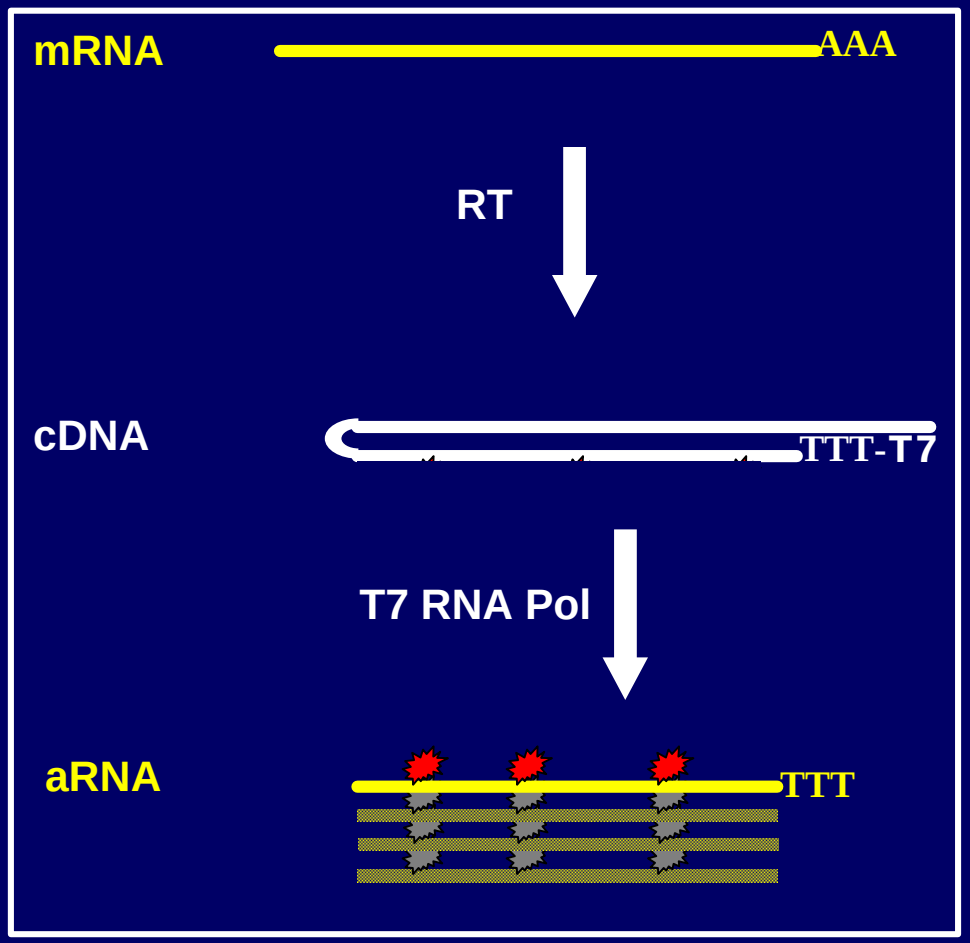
Ex situ Oligo Gene Expression Microarrays

- A “home grown” format distinct from any commercial platform.
- 45mer oligonucleotides spotted onto slides
- Covalent attachment to surface, 5’ anchored
- Hybridizations use cDNAs labeled with Cy3 & Cy5. aRNA labeled with ALEXA/Cy dyes also worked well.
- We evaluated an array of:
  300 genes x 6 oligos/gene = 1800 oligos
  x 4 replicate spots per oligo = 7200 spots

Merck Custom Oligonucleotide Microarrays
Labeling Hybridization Sample:

- cDNA-Cy3/Cy5
- amplify aRNA-ALEXA488/546

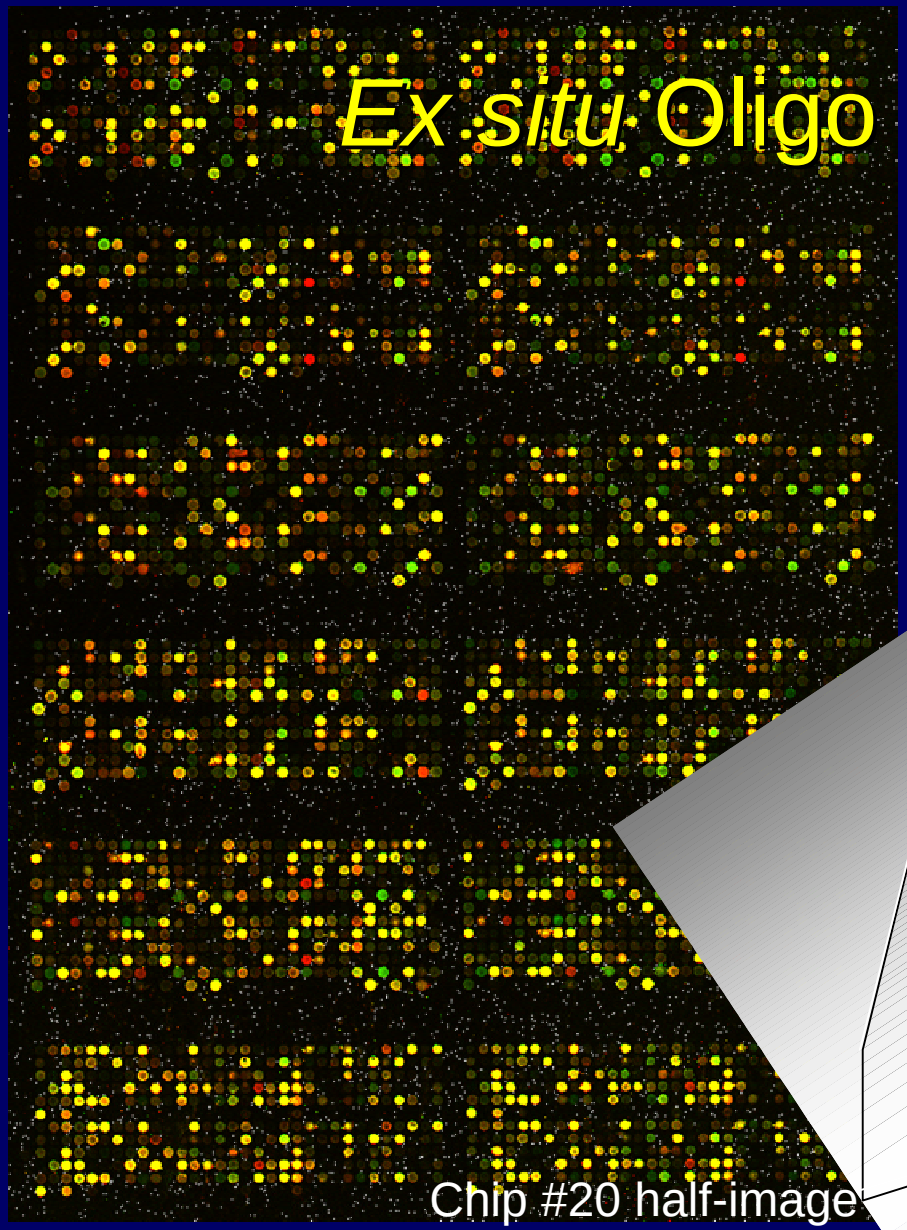
"DNA & Protein are RNA's way of making more RNA"



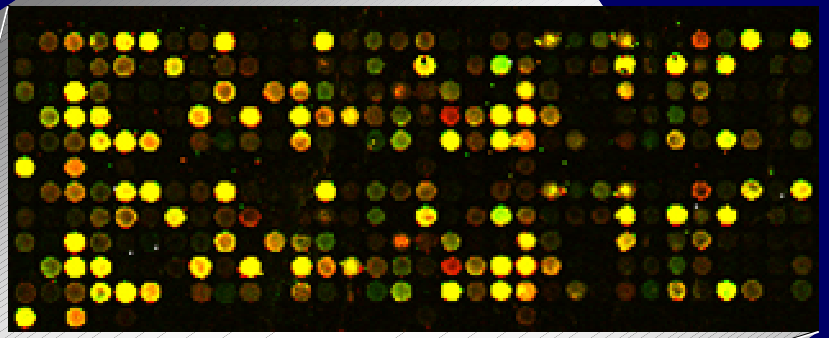
D. Gerhold a

Ex situ Oligo Microarrays

- **Quality of arraying and hybridization is excellent.**
- **Display:**
  - **red=drug-induced gene**
  - **green=drug-suppressed gene**



Chip #20 half-image

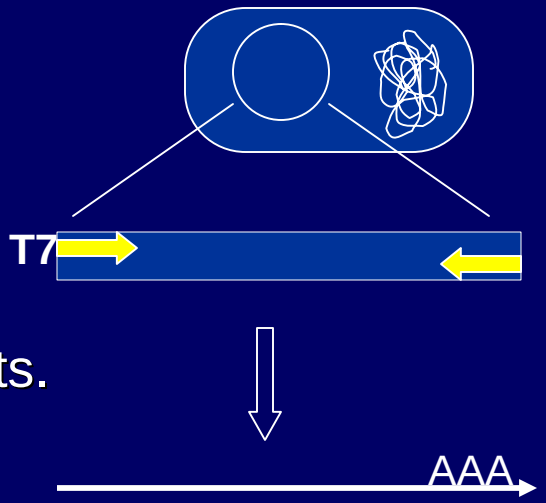


D. Gerhold

Public disclosed

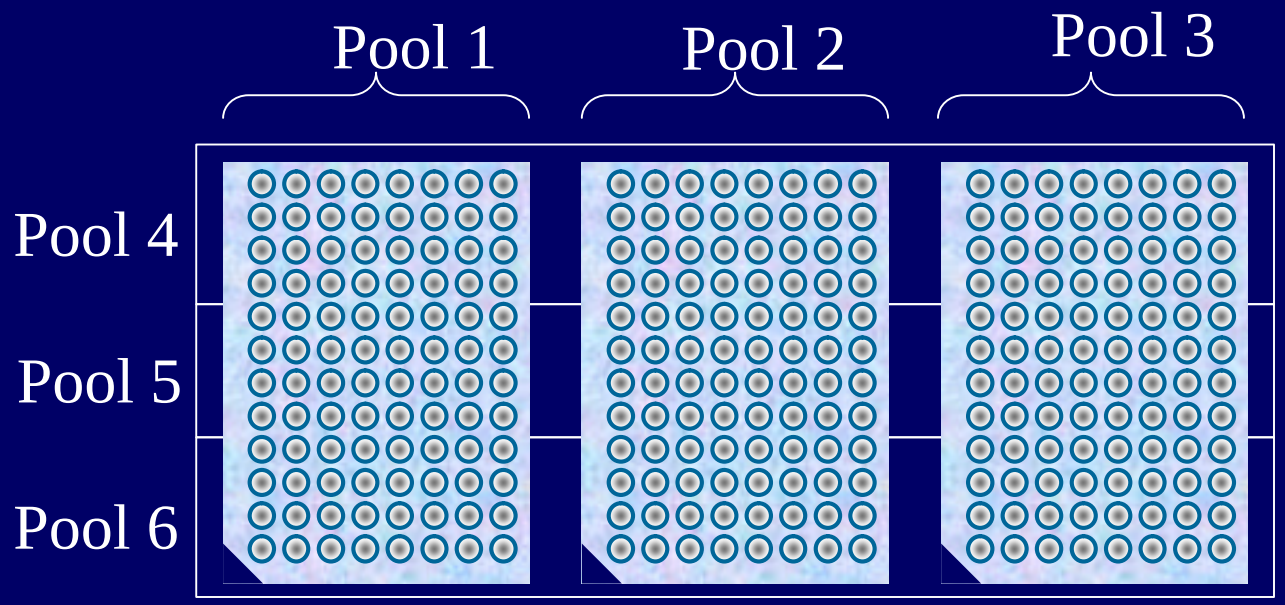
Spiking Strategy to Assess Chip Performance

- Buy EST clones for all arrayed genes.
- Use PCR to amplify insert DNA.
- Use T7 RNA pol to make *in vitro* transcripts.
- Quantitate & pool subsets of transcripts.
- Spike transcript pools into yeast total RNA
- cDNA was labeled with Cy3 & Cy5 by reverse transcription .



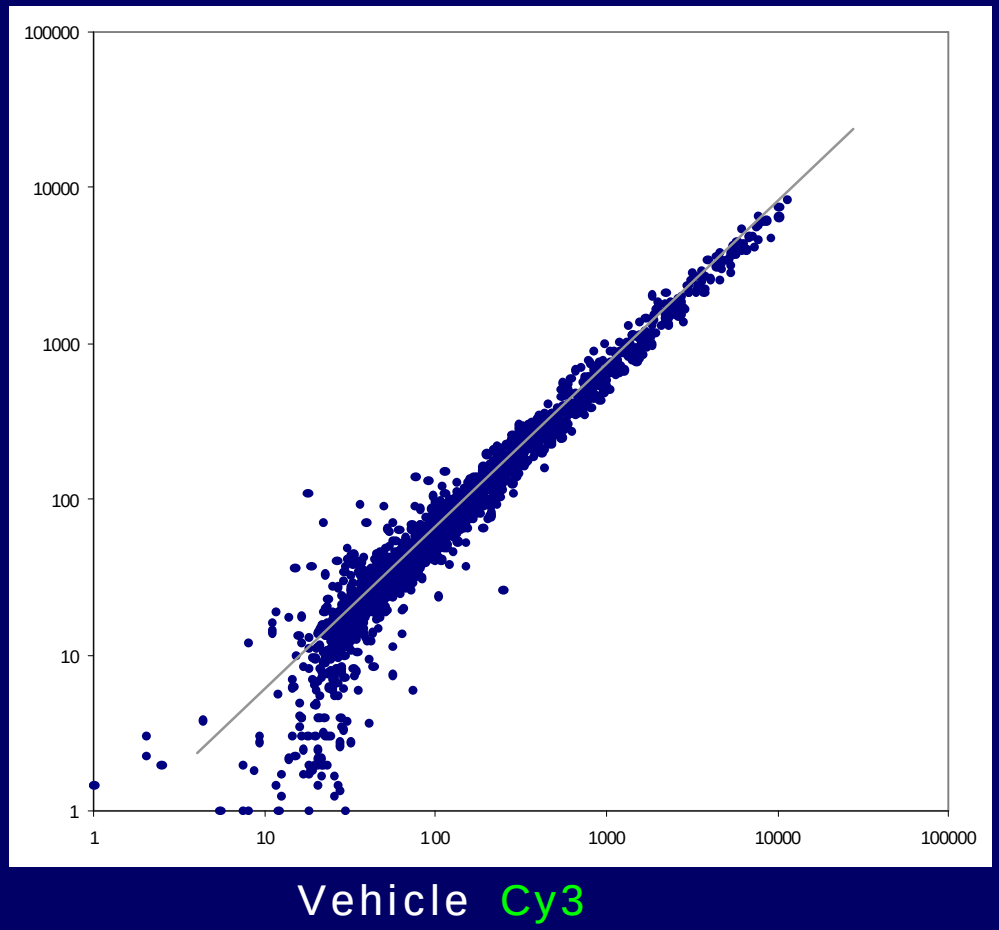
Transcript Pooling Strategy

- 3 x 96 EST transcripts represent genes on the chip.
- Each pool of 96 transcripts is hybridized to a chip at each of 4 concentrations 1/5,000 - 1/135,000.
- Each gene-transcript is separated from 88% of other genes.



D. Gerhold a

Reproducibility

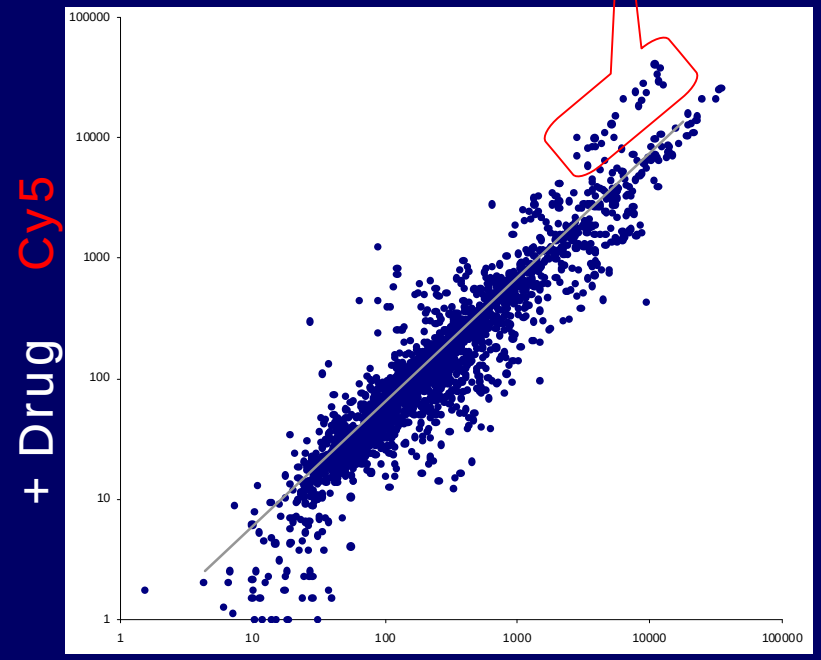


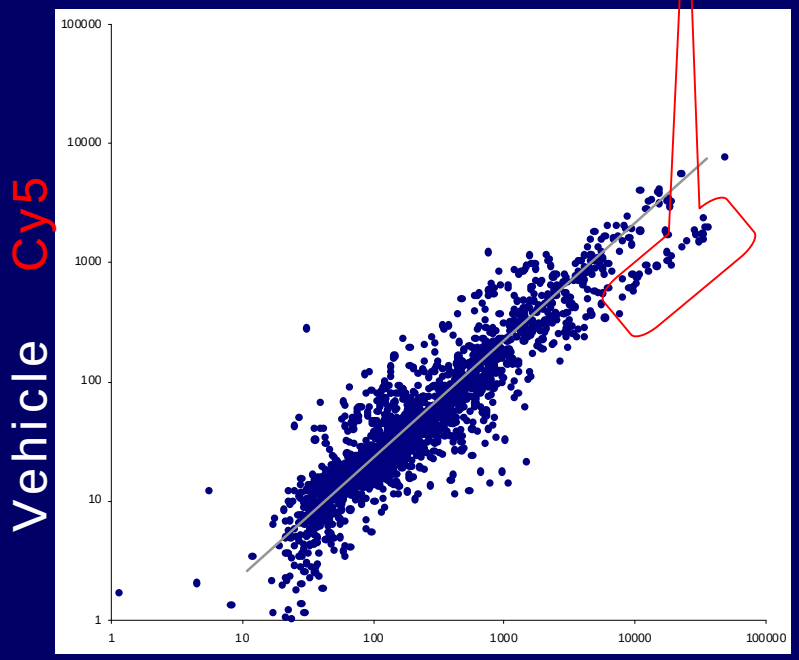
- **Compared two aliquots of same RNA sample**
- **For individual Oligos c.v.$_{>100}$ = 17%**

D. Gerhold a

Two-Color Hybridization: Peroxisome Proliferator-Treated Mouse Liver

Same genes/oligos occur in both "spurs", including classical responses: Acyl-CoA Oxidase, CYP4A10, and Carnitine Palmitoyl Transferase 1

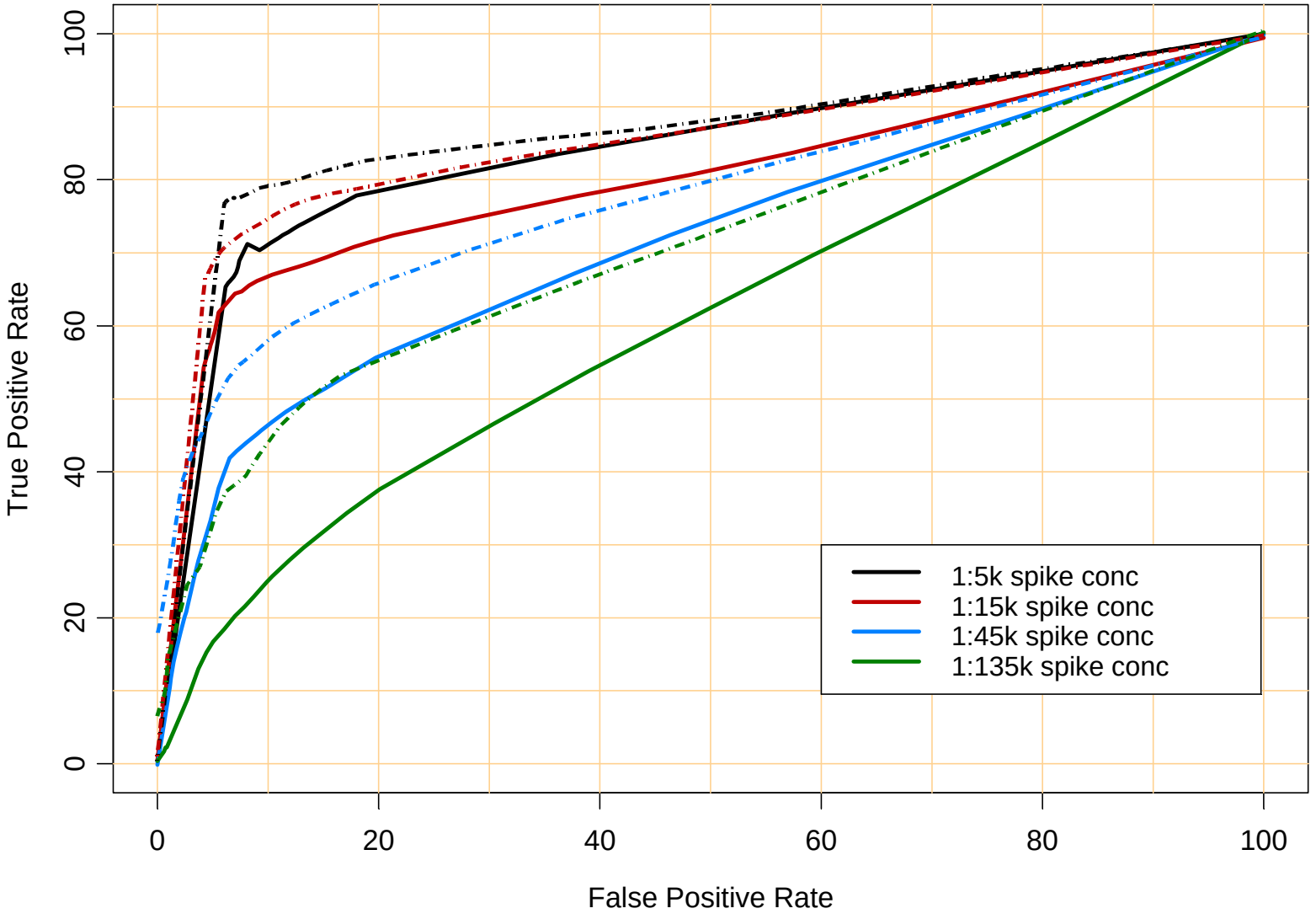




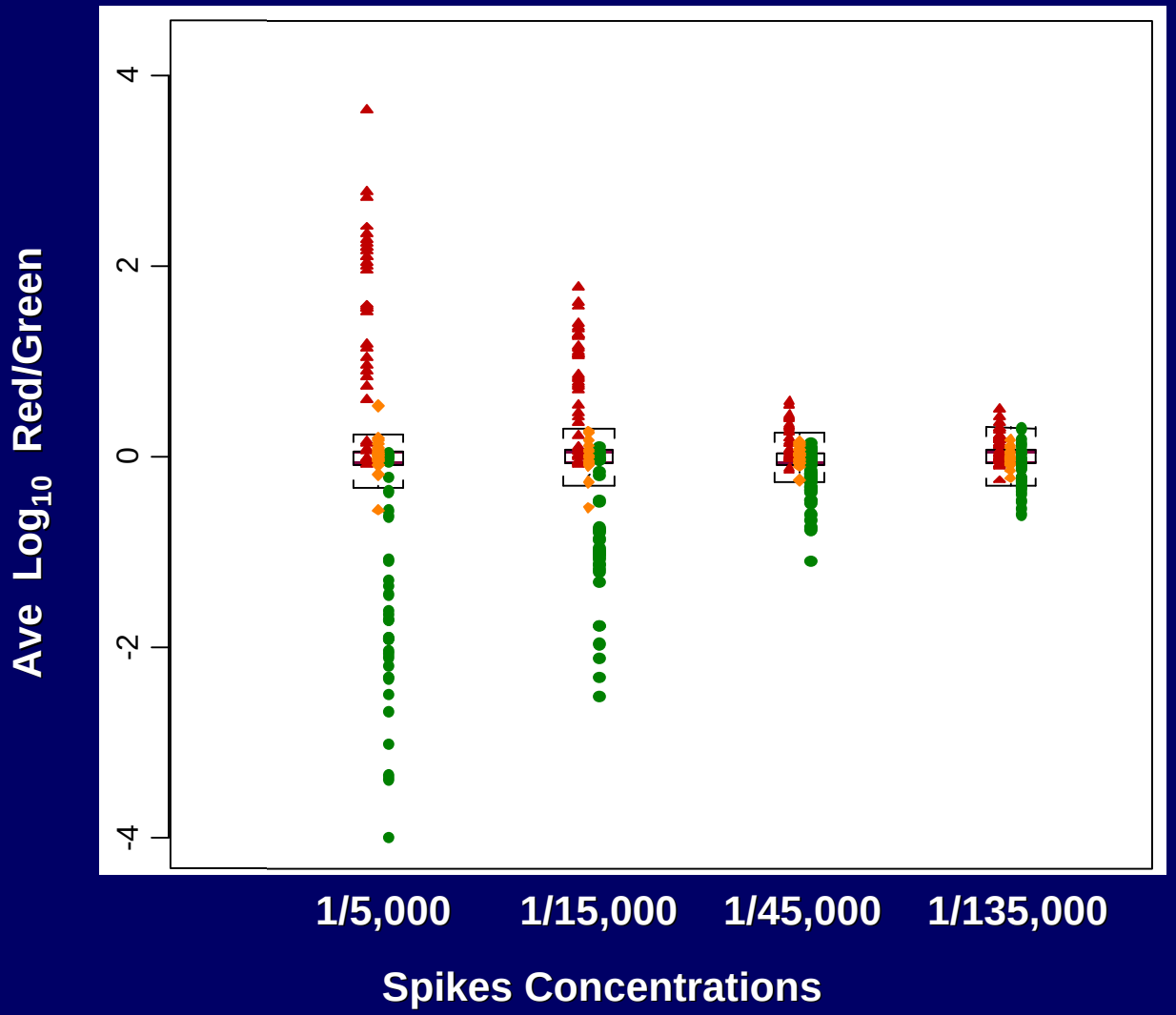
D. Gerhold a

Overall True Positive vs Flase Positive Rate by Spike Concentrations

Solid line = Individual result from diff chip, Doted line = Using results from the best oligo for spiked genes

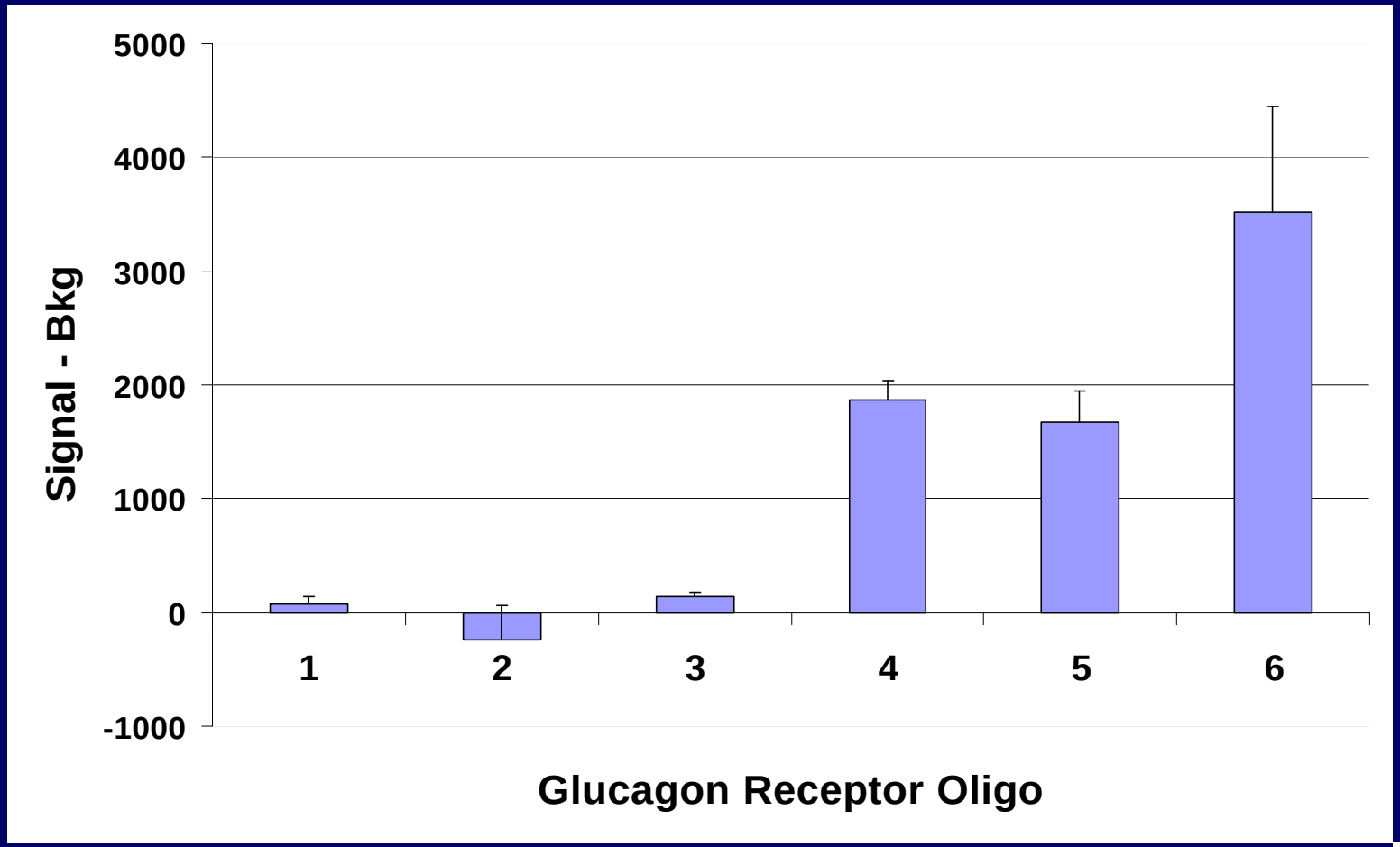


Evaluating sensitivity of probes
Pooled Spikes Experiment: Groups 2 and 5



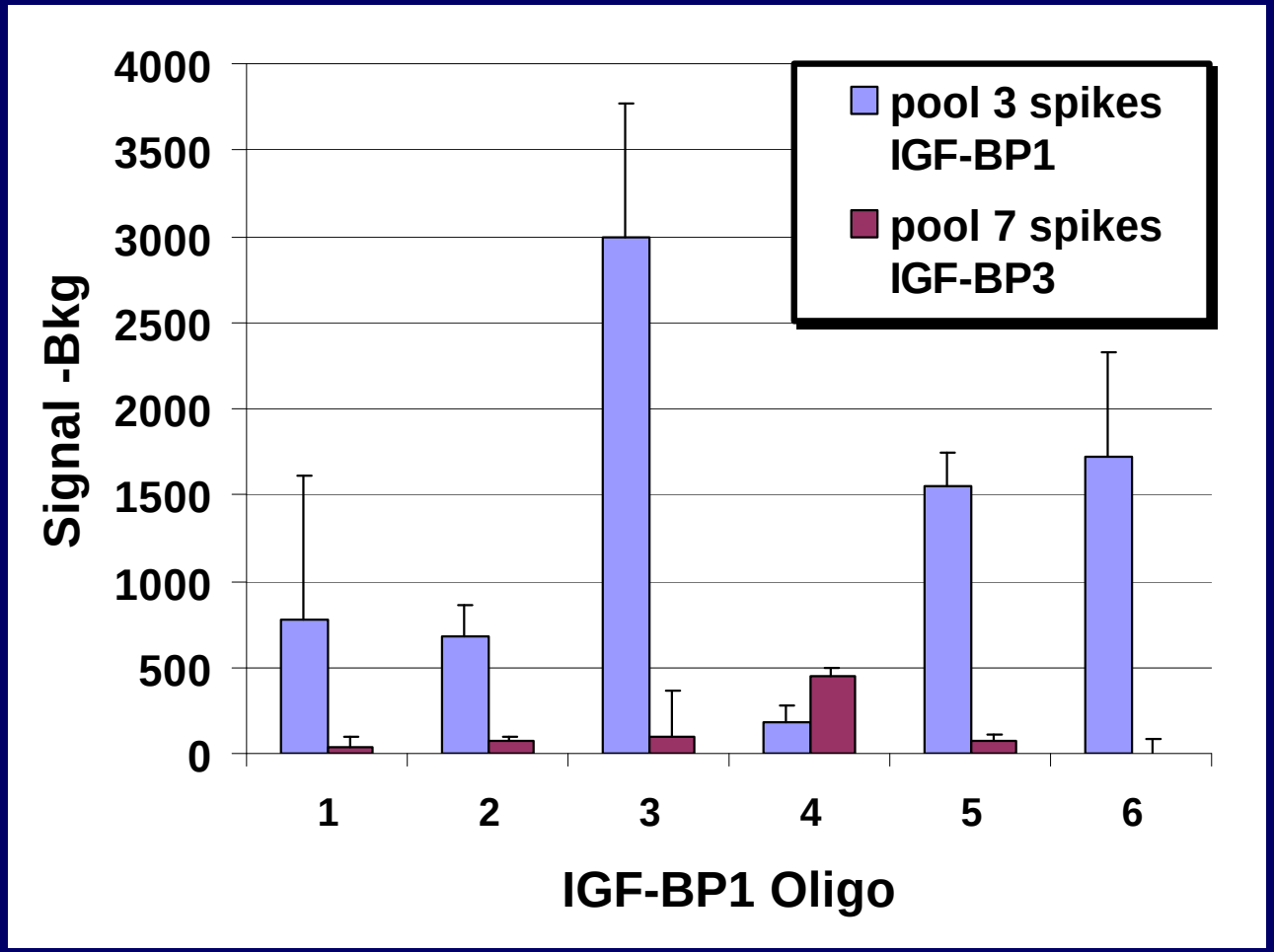
D. Gerhold a

Sensitivity Example: Glucagon Receptor Oligos



D. Gerhold a

Specificity Example: IGF-BP1 Oligos



D. Gerhold a

Why Do Transcript Pooling "TraP"??

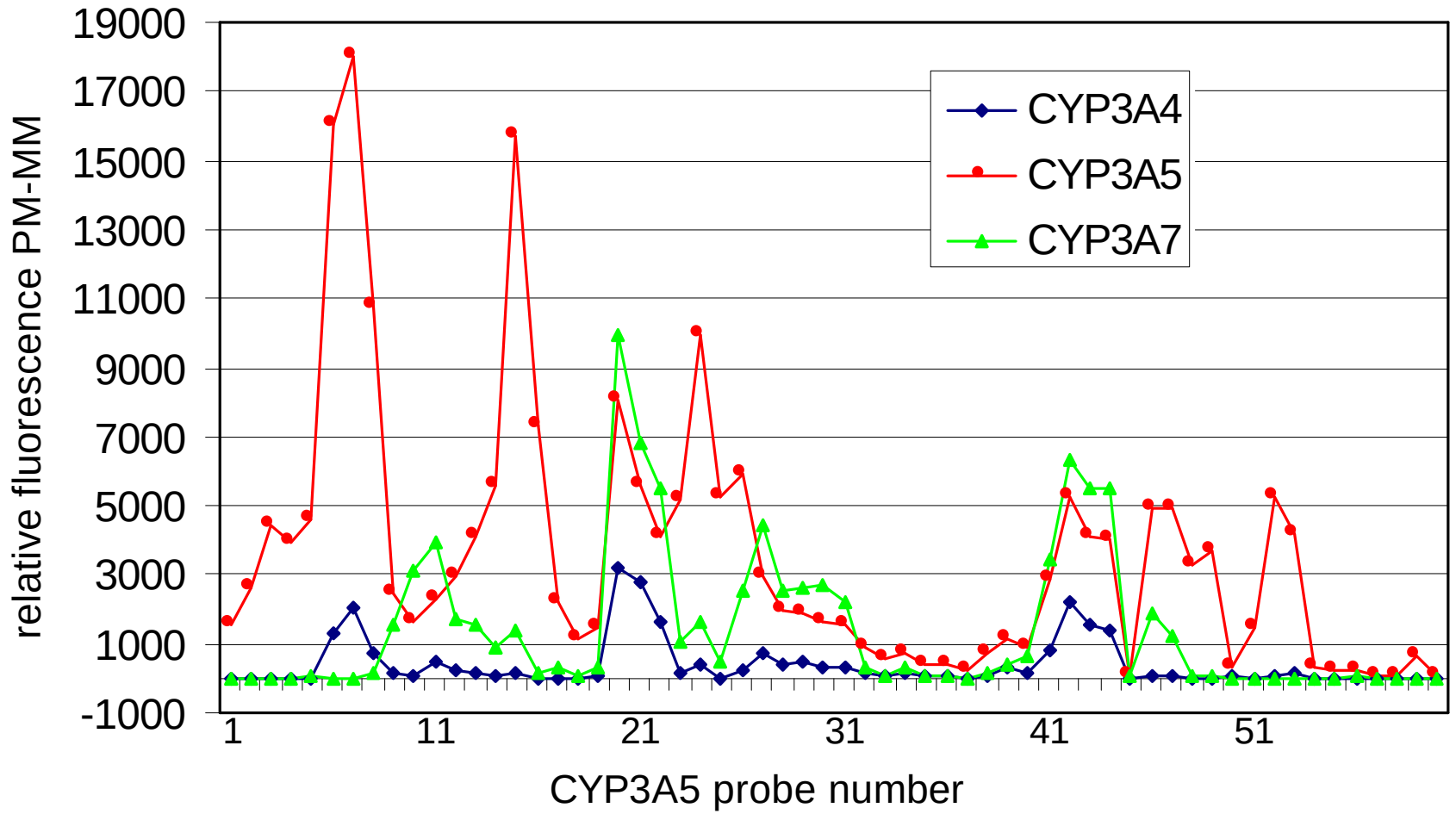
- Sensitivity
  - Determine the detection limit for each gene.
  - Identify which oligos are most sensitive for each gene.
- Specificity
  - Determine which oligos are gene-specific.
  - Determine which oligos yield lowest "noise" for each gene.
- Reproducibility
  - How many replicate hybridizations do I need to detect gene regulation events?
  - How many replicate hybridizations do I need to eliminate false-positive events?
- Transcript pooling need be done only once per organism. These RNA transcripts could be made available by a central organization that collects genes/clones.

Analysis of Hybridization Specificity in the Human CYP Gene Superfamily

	Spiked *CYP3A* gene:		
Chip Probes:	***CYP3A4***	***CYP3A5***	***CYP3A7***
CYP2A6	**2938**	**2892**	**3153**
CYP2A13	1	0	0
CYP2C8	0	0	0
CYP2C9	9	8	8
CYP2C18	2	0	1
CYP2C19	**1800**	**2227**	**2199**
CYP2D6	**2901**	**2518**	**2284**
CYP3A4	**491**	11	15
CYP3A5	270	**2904**	1140
CYP3A7	8	5	**3467**

D. Gerhold a

Analysis of Hybridization Specificity in the Human CYP Gene Superfamily

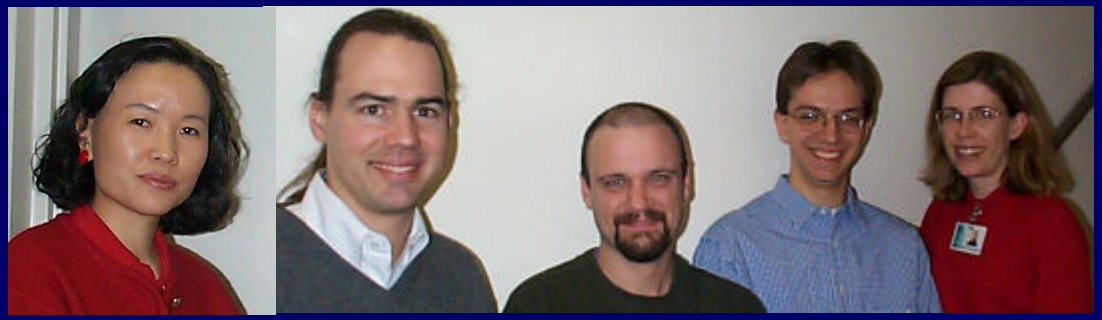


D. Gerhold a

Platform Performance Evaluation: *Logistics of gene/transcript collections*

- Collect 1,000s of genes for human, mouse, rat, etc. as plasmid clones in *E. coli*. NIST or a commercial vendor?
- A diverse gene superfamily should be represented completely for each organism to assess gene specificity (*e.g. CYP*'s).
- Generate, purify, quantitate RNA from each gene. A minimum of ~1,000 nt needed at 3' end, complete genes preferred. Repeat this process ~annually(?) at a central facility. NIST?
- Generate "normalized" pools of RNAs, (*e.g.* 16 pools) to send to requesting labs.

Collaborators: *To steal ideas from one person is plagiarism; to steal from many is research.* A. Felson



DNA Arrays:

Hong Jin, Robert Phillips, Chris Wright, Andrew Kernytsky, Allyson Cole-Strauss

Jian Xu, Meiqing Lu, Bob Benz, David Gerhold

Oligo Database:

Gary Mallow, Jeannette Cabardo

Biometrics:

Lisa Ying, Vladimir Svetnik, Dan Holder

Drug Metabolism:

Tom Rushmore

Merck Oligo Design "M*OD*" Automates Oligo Selection

- Batch input of sequences, uses PERL "shell" around *PRIMER3* from Whitehead Institute, iterative in relaxing selected parameters.
- Masks repeats in gene sequence, runs of 1 base, hairpins.
- Optimizes deltaG, GC-content, 3' 7 base sequence,
- Picks *e.g.* 6 oligos per gene with <50% overlap.
- Oligos triaged to distinguish related genes.
- Output oligos tracked in database "*BR-OLIGO" from* design-to-freezer-to-array using barcodes.

“Omics” Nobel Prize

Stanford Report, October 10, 2001

Stanford's Spence wins Nobel for Econ-omics

BY BARBARA BUELL

A. Michael Spence, Philip H. Knight Professor, Emeritus, and former Dean of the Graduate School of Business, was awarded the 2001 Nobel Memorial Prize in Economic Sciences Wednesday. He shares the prestigious $1 million prize with George A. Akerlof of the University of California at Berkeley and Joseph E. Stiglitz of Columbia University.

May's Law

The quality of correlation is inversely proportional to the density of control. (The fewer the data points, the smoother the curves.)